A TWO TRANSISTOR FLASH MEMORY CELL FOR USE IN EEPROM ARRAYS WITH A PROGRAMMABLE LOGIC DEVICE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to semiconductor memory and in particular to a two transistor flash EEPROM cell for use in low-voltage, low-power, high-speed, and high-density applications, such as complex programmable logic devices.

2. Description of the Related Art

Prior art Flash memory technology used in applications for devices, such as PLD (programmable Logic Device), PAL (programmable array logic) and EPLD (erasable programmable logic device), are two transistor memory cells comprised of an NMOS nonvolatile memory device and an NMOS access device. These two transistors are connected in series to form the basic flash memory cell for programmable logic devices. The NMOS nonvolatile memory device of prior art is an asymmetric device with respect to the source and drain as well as operating conditions. A high voltage is required across the drain and source region during programming which requires a longer channel gate length to prevent punch through. This causes a physical limitation on how small the cell can be made and in turn limits the use of the cell in ultra high integration levels of the flash memory below 0.18um technology.

US 6,108,239 (Sekariapuram et al.) is directed toward a compact nonvolatile

relative to the surface of the semiconductor substrate. In US 6,078,521 (Madurawe et

al.) a nonvolatile memory cell is directed toward a compact layout and a high logic output voltage. In US 5,914,904 (Sansbury) is directed toward a nonvolatile memory cell that has a read device, a program device and a tunnel diode. In US 5,904,524 (Smolen) a device and method is directed toward an EEPROM device that has a self aligned tunnel window with low gate capacitance and avoids defects caused by field oxide induced stress in the tunnel oxide. In US 5,914,514 (Dejenfelt et al.) a flash EEPROM cell is directed toward a two transistor cell for high speed and high density PLD applications. The storage transistor is directed toward preventing problems with over erase and punch through, and allows scaling the gate length to allow 5V cell programming.

In US 5,862,082 (Dejenfelt et al.) a device is directed toward a flash EEPROM cell that has two transistors with one transistor being a floating gate type device with asymmetric source and drain. In FIG. 1 is shown a cross sectional view of prior art for aflash EEPROM cell 100 similar to that of US 5,862,082. The flash EEPROM cell includes a nonvolatile memory transistor 101 and an access transistor 102. The nonvolatile memory transistor 101 is fabricated within a p-well 103 and is a stack type double poly transistor, which includes a thin tunnel oxide film 108, a floating gate 109, an interpoly dielectric layer 110, a control gate 113, n+ source region 123, n+ source/drain region 122 and n- type region 124. The access transistor 102 is fabricated within a p-well 103, which includes a source/drain region 122, n+ drain region 121, a gate dielectric layer 111 and an access gate 112. The P-well 103 is formed within an n-

104 creating a contact region for the N-well. A P+ type region 107 is formed in p-well 103 creating a contact region for the P-well.

FIG. 2 of the prior art is a circuit diagram illustrated as a 2x2 array 600, formed by a plurality of identical flash EEPROM cells 601, 602, 603 and 604, which are comprised of the flash EEPROM cell 100 . Each column of cells has separate source lines 631 and 641 and drain bit lines 632 and 642 for high-speed PLD applications. Lines 611 and 621 connect to the control gates of the storage transistors of cells 601and 602, and cells 603 and 604 respectively. Lines 612 and 622 connect to the access gates of the access transistors of cells 601 and 602, and cells 603 and 604 respectively.

FIG.3 is a plane view of prior art of the flash EEPROM cells 601 to 604 of array 600, and corresponding to the circuit diagram in FIG. 2. Control gate lines 611 and 621 run horizontal across the plane view of the cells connecting to the control gates of the storage transistors of cells 601 and 602, and cells 603 and 604 respectively. Similarly access lines 612 and 622 run horizontally across the plane view of the cells connecting to the gates of the access transistors of cells 601 and 602, and cells 603 and 604 respectively. Source lines 631 and 641 run vertically across the plane view connecting to the sources of the storage transistors of the EEPROM cells 601 and 603, and cells 602 and 604 respectively. Similarly bit lines 632 and 642 run vertically across the plane view connecting to the sources of the storage transistors of the EEPROM cells 601 and 603, and cells 602 and 604 respectively.

operations use the Fowler-Nordheim tunneling. The electrons are injected into the floating gate of the storage transistor of the cells by channel-erase operation which increases the threshold voltage (Vt) of the storage transistor. The electrons are extracted out of the floating gate of the storage transistor by an edge-program operation which decreases Vt. The erase operation is performed in a blanket mode. All the cells 601-604 in array 600 are erased simultaneously. A high voltage of 8 to 10 volts is applied to the control gate word lines 611 and 621, and a negative high voltage of -8 to -10 volts is applied to source lines 631 and 641. At the same time, the same negative high voltage -8 to -10 volts is applied to p-well 103, and Vcc supply voltage of 3.3 volts is applied to n-well104. Thus the threshold voltage Vt of the storage transistor of the cells is increased for the erase operation.

Continuing to refer to FIG. 4, a program operation is performed bit by bit on each word line. For example, cell 601 is programmed and cell 602 is program inhibited by applying a voltage of -7 to -11 volts to the control gate word line 611 and applying a voltage of 8 volts is to the access gate 612. The source lines 631 and 641 are maintained at a high impedance state, and a voltage of 5 to 8 volts is applied to the bit line 632 with 0 volt applied to the bit line 642. P-well 103 is maintained at 0V and the n-well 104 is maintained at the Vcc supply voltage. 3.3 volts. As a result, cells 601 and 602 are placed in the program and program inhibition mode. respectively.

Continuing to refer to FIG. 4, the two-transistor flash EEPROM cell of prior art created with an n+ source/drain region 122 and an n- type region 124 that are made to

operation. For programming purposes, it should be noted that there is a reverse bias voltage between n- type region 124 and p-well103. As a result of the negative high voltage applied to the gate of the memory cell, electron-hole pairs will be generated and holes are accelerated onto the floating gate under the high electrical field. A certain amount of holes will be trapped in the tunnel oxide, and will degrade the oxide after cycling.

The memory cell of prior art is made non-symmetrical with respect to source and drain junctions in terms of cell structure and operating conditions. A high voltage of more than 5V across lightly doped N implant drain region and source region during program operation is required. Therefore, the prior art cell of FIG. 1 through FIG. 4 needs a longer channel gate length to prevent punch-through. Otherwise, the impact of punch-through is the degraded efficiency in program operation. The prior art has limitations in shrinking the length of the cell for ultra-high integrated flash memory below 0.18um technology.

SUMMARY OF THE INVENTION

An objective of the present invention is to use a fully symmetrical, scaleable Flash EEPROM memory cell having a storage transistor coupled in series with an access transistor, and which may be used to form compact flash arrays for

Another objective is to replace the edge-program operations in prior art by channel program operation to remove the high voltage drop across the channel region of the storage transistor of the cell. Such removal of the voltage gradient across source and drain permits the use of a shorter channel length for high-density applications, such as in complex programmable logic devices (CPLDs);

Still another objective of the invention is to provide channel erase by reducing the threshold of flash EEPROM cell with immunity from the over erase problem;

Yet another objective of the invention is to provide a preferred voltage for non-selected word lines for a channel program operation so that the disturbance of the Vt of non-selected cells can be eliminated or substantially reduced;

Yet another further objective of the invention is to provide the preferred voltages for word lines, bit lines, source lines and P-wells so that the aforesaid drawbacks of asymmetrical cell can be eliminated and replaced by a highly scaleable symmetrical cell;

Still another objective of the invention is to provide the preferred voltages for word lines, bit lines, source lines and P-wells so that the aforesaid drawback of channel punch-through can be eliminated and high current program and erase operations can be reduced:

Still another further objective of the invention is to provide the preferred voltages for word lines, bit lines, source lines and P-well so that the aforesaid drawback of low cycling can be improved with high endurance.

Still yet another objective of the invention is to provide a two-transistor cell with over erase immunity for a simplified on chip state machine without program and erase verifications;

Still yet another objective of the invention is to provide a two-transistor cell with a floating gate device fully compatible with the access device and any peripheral NMOS single poly devices in terms of process step and device structure.

The present invention provides a highly scalable, two-transistor, Flash EEPROM cell with a fully symmetrical source and drain structure. The two-transistor cells are comprised of one NMOS floating-gate nonvolatile memory (NVM) and one NMOS access device. These two transistors are connected in series to form a flash cell in accordance with the present invention. A plurality of these two-transistor cells are used to form a matrix array with a plurality of columns and rows. The columns are comprised of metal bit lines and metal source lines coupling to drains and sources of the two-transistor cells. The bit lines and source lines are decoded by a bit line decoder and a source decoder, respectively, and run vertically in parallel through flash cell array. The rows are comprised of a plurality of word lines and access lines. The word lines are connected to the gates of NVM devices of the two-transistor cells, and the access lines are connected to the gates of access devices of two-transistor cells. The word lines and access lines run horizontally across the flash cell array. The bit lines are connected to sense amplifiers and source lines are connected to source line decoders. The word lines and the access lines are driven by X-decoder and access decoders. The preferred

high-density programmable logic devices, comprising PLD, PAL, and EPLD applications.

Unlike the prior art flash technology used in PLD, PAL and EPLD, the two-transistor flash cell utilizes the preferred Fowler-Nordheim (FN) channel erase and FN channel program methods to allow a fully symmetrical cell structure with ultra-high cell scalability and ultra-low program and erase current. With the two-transistor cell structure, the cell of the present invention provides a solution to fully eliminate the conventional over erase problem that occurs in one transistor flash cells. This is because the access transistor will not conduct current even though the NMOS floating-gate nonvolatile memory is over erased. In addition, the flash cell of the present invention can be made to avoid punch through and disturbance problems particularly in sub-micron flash technology by using the preferred channel erase and channel program operations.

The main advantage of the channel erase and the channel program is the elimination of the high voltage drop across the drain and source of the flash cells during program and erase operations. As a result, the channel length of the flash cell is no longer limited by erase and program operations and is only determined by the read operation. Conventionally, read operations in flash memory is designed to operate with less than 1V across the drain and source nodes; therefore, the length of the flash cell of the present invention can be easily shrunken producing a high cell scalability. Also a large read current can be achieved in accordance with the present invention.

state-machine can be accomplished to control the reduced on-chip operations for the applications requiring in-system re-programmability. According to the present invention, the charge storage gate electrode (floating gate electrode) of the nonvolatile cell is formed on the surface of an active region by means of a polysilicon layer on top of a first insulating film. A control gate electrode is formed on the surface of the charge storage gate electrode by means of a polysilicon layer on top of a second insulating film. The bulk of the cell can be either formed on P-substrate or a P-well within a deep N-well on P-substrate.

BRIEF DESCRIPTION OF THE DRAWINGS

This invention will be described with reference to the accompanying drawings, wherein:

FIG. 1 is a cross sectional view of flash EEPROM cell of prior art containing an access device;

FIG. 2 is a circuit diagram of prior art of a 2x2 flash EEPROM cell array;

FIG. 3 is a plan view of a 2x2 flash EEPROM cell array of prior art with two separate metal source lines and two bit lines;

FIG. 4 is a table summarizing the voltages for read, erase, program and program inhibit for a flash EEPROM cell array of prior art;

FIG. 5 is a cross sectional view of a two-transistor flash EEPROM cell made on

FIG. 6 is a circuit diagram of a 2x2 flash EEPROM cell array in accordance with one embodiment of the present invention;

FIG. 7 is a plan view of a flash EEPROM cell array in accordance with one embodiment of the present invention; and

FIG. 8a and 8b summarize the preferred bias conditions for read, erase, program and program inhibit modes of a flash EEPROM cell array in accordance with two embodiments of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In FIG. 5 is shown a cross sectional view of a two transistor flash EEPROM cell 300 in accordance with one embodiment of this invention. The storage transistor of the EEPROM cell 300 containing a floating gate 11 and a control gate 10 has an n+ source 22 and n+ source/drain 13 that is fully symmetrical on p-well 140. The P-well 40 is formed within a deep n-well 41 on a p-substrate 16. The EEPROM nonvolatile memory storage device is a stack type double poly transistor, which includes a thin tunnel oxide film 15, a floating gate 11, an interpoly dielectric layer 14 and a control gate 10. The flash EEPROM cell 300 of the present invention is preferably programmed and erased by a Fowler-Nordheim (FN) channel erase 38 and FN channel program 39 as indicated by the two arrows. In the prior art, channel erase is used to increase the Vt of cells but

channel erase 38 is used to decrease the Vt of the cells and channel program 39 is used to increase the Vt of the cells.

Continuing to refer to FIG. 5, during a channel program operation, no reverse biased voltage between n+ source/drain region 13 and the p-well 40 occurs, as with an edge-program in the prior art , and there is no voltage drop across the channel region between source and drain. With a positive high voltage being applied between the control gate 10 of the storage transistor and the n+ source/drain region 13, n+ source region 22 and the p-well 40, an electric field with sufficient magnitude and polarity will result in FN tunneling. . Holes are expelled to p-well 40 and electrons are attracted to floating gate 11 No holes will be trapped in the tunnel oxide as with the prior art. As a result, a better endurance for program and erase cycles can be achieved.

Continuing to refer to FIG. 5, since the cell structure of the present invention uses both FN Channel-erase 38 and FN channel-program 39 schemes, the N- region 124 of prior art shown in FIG. 1, which is used for edge-program, is eliminated. Therefore, junctions of drain 13 and source 22 shown in FIG. 5 can be made fully symmetrical allowing a very small channel length. One preferred method of forming a floating gate device comprising the control gate 10 and the access device comprising the access gate is to make both fully process compatible. This can be done by making the tunnel oxide 15 under the floating gate 11 and the gate oxide 17 under the access gate 18 have the same oxide thickness and the same material. Similarly, the thickness of poly floating gate 11 and poly access gate 18 are made to be the same. The source and

time with the same process step. The second gate 10 is formed after the completion of access gate 18 and floating gate 11. Thus, the flash cell of the present invention uses the preferred channel erase 38 and channel program 39 methods.

FIG. 6 is a circuit diagram illustrated as a 2x2 array 500, formed by a plurality of identical two-transistor flash EEPROM cells 501, 502, 503 and 504 as shown in FIG. 5 as cell 300. Each column of cells has separate metal source bit lines 522 and 532, and metal drain bit lines 521and 531 for high-speed PLD applications. The control gate lines, 512and 513, are coupled to flash cells 501 to 504. The access-gate lines, 511 and 514, are coupled to single poly cells 501 to 504.

FIG.7 is a plan view of a 2x2 cell array 500 comprising flash EEPROM cells 501 to 504, control gate lines 512 and 513,access gate lines 511 and 514, metal drain bit lines 521and 531, and metal source bit lines of 522 and 532 according to FIG. 6. The control gate of cell 501 (corresponding to control gate 10 in FIG. 5) is connected to the control gate of cell 502 along a horizontal direction, thereby forming a first control gate word line 512. Similarly, the control gate of cell 503 is connected to the control gate of cell 504 along a horizontal direction, thereby forming a second control gate word line 513. The access gate of cell 501 (corresponding to access gate 18 in FIG. 5) is connected to the access gate of cell 502 along a horizontal direction, thereby forming a first access gate word line 511. Similarly, the access gate of cell 503 is connected to the access gate of cell 504 along a horizontal direction, thereby forming a second access gate word line 514. The source region of the flash EEPROM cells 501 and 503

of flash EEPROM cells 501 and 503(corresponding to drain region 19 in FIG. 5) are connected by a metal line along the vertical direction, thereby forming a first drain bit line 521. In a same manner, the second source bit line 532 is formed by a vertical metal line connecting the source regions of cells 502 and 504, and the second drain bit line 531 is formed by a vertical metal line connecting the drain regions of cells 502 and 504.

Continuing to refer to FIG 7, it should be noted that field oxide is present in all regions except for those regions defined by drain region 19, source region 22, source/drain region 13 and the channel region of cells 501-504; therefore, the source bit lines 522 and 532 are segmented and each is separately interconnected by metal through contacts. Similarly, the drain bit lines 521 and 531 are segmented and each is separately interconnected by metal through contacts. The channel region of cells 501-504 is formed to be in parallel with the direction of source and drain bit lines.

FIG. 8a shows the preferred bias conditions for control gate word lines, access gate word lines, drain bit lines and source bit lines for the array 500 shown in FIG. 6 and 7 assuming the cell is fabricated in a p-well within a deep n-well on a p-substrate. Voltage conditions for erase, program, program-inhibit and read operations are shown . Referring to FIG.8a, an erase operation is performed with all the control gate word lines set at -10V, all of the source bit lines and the p-well at +5V, all of the access gate word lines at 0V and the drain bit lines in high impedance. It should be noted that the -10V and +5V are exemplary values, and the exact value and time of all nodes voltage are subject to different flash technologies. A voltage of -15V across control gate and p-well

(on-state) after a predetermined erase time. The FN channel erase operation is preferably performed in a blanket mode in a PLD application.

Continuing to refer to FIG 8a, in the channel program mode, the selected control gate word line is coupled with Vpgm, the access gate word line with Vcc, the drain bit line and p-well with -5V and the source bit lines are in high impedance, performs FN program operation. The non-selected control gate word lines are tied to -2.5V and the non-selected access gate word lines are tied to Vcc to minimize the build-up of disturbance between the gate and the p-well and between the the gate and source/drain. Vpgm is an adjustable voltage input for the control gate of the cells to meet different Vt requirements in program operation. Vpgm may vary from as low as 7V up to about 10V or beyond, and is a ramping voltage that takes 0.5V steps between 7V and 10V. It should be noted that all the voltages are exemplary values for better understanding of present invention. The exact value and time of all nodes in this operation varies with different flash technologies.

Continuing to refer to FIG 8a, to inhibit programming of selected cells on the selected word line during the program mode, a voltage Vd of 0V is applied to the drain bit line of unselected cells. Compared with the voltage of selected programmed cells, a 5V reduction of the electric field will prohibit the FN tunneling current from occurring in the selected non-programmed cells. As a result, the program inhibited cell will not be selectively programmed during the time of the programming operation.

Continuing to refer to FIG. 8a, to read a cell the drain bit lines are coupled with

0V. These voltages are exemplary values to allow better understanding of the present invention. Within a PLD, multiple word lines of memory cells can be simultaneously are coupled with multiple logic inputs.

FIG. 8b shows the preferred bias conditions for control gate word lines, access gate word lines, drain bit lines and source bit lines for the array 500 shown in FIG. 6 and 7 assuming the cell is fabricated on a p-substrate. Voltage conditions for erase, program, program-inhibit and read operations are shown. Referring to FIG. 8b, the speed of the sense amplifier design is required to satisfy two extreme conditions, one memory cell conducting current and all memory cells conducting current. When all the memory cells are conducting, the voltage drop in the selected bit lines can become significant. The time for bit line voltage recovery depends upon the loading of the bit line and cell current. In order to meet high-speed of PLD applications, the two-transistor flash EEPROM cell must have a high transconductance. Shown in FIG. 8b, is a channel erase operation where all word lines are coupled to -15V, all source bit lines and the p-substrate are coupled to 0V and the drain bit lines are in high impedance. The control gate voltage of -15V is an exemplary value. The exact value and time of all nodes voltage are subject to different flash technologies. A voltage of -15V voltage across the gate and p-substrate will result in high electric field in channel region of the cells. The tunneling electric field will attract electrons flowing from floating gate to p-substrate to decrease the Vt (on-state) of the cells after a predetermined erase time. The erase operation will be performed in a blanket mode in the PLD application of the present

Continuing to refer to FIG. 8b, in the channel-program mode the selected control gate word line is coupled to Vpgm, the access gate word line is coupled to 8V, the drain bit line and p-substrate are coupled to 0V, and the source bit lines are in high impedance. The Vpgm is an adjustable voltage input for the control gate of the cells to allow different Vt requirements in the program operation. Vpgm may vary from as low as 12V up to about 15V or beyond, and is a ramping voltage that takes 0.5V steps between 12V and 15V. It should be noted that all the voltages are exemplary values for better understanding the idea of the present invention. The exact value and time of all nodes in program operation are subject to the different requirements of different flash technologies.

Continuing to refer to FIG. 8b, to inhibit programming on the selected cells in the selected word line during the program mode, a drain voltage (Vd) of 5V is applied to the selected drain bit line. The voltage drop across the channel region of the cell is greatly reduced, and the FN program operation is inhibited. The preferred program inhibit conditions reduce the tunneling current in the selected cell by at least two orders of the magnitude as compared with the selected programmed cells. As a result, the Vt changes in the program inhibited cells are negligible during the time of the programming operation. In a read operation the drain bit lines are coupled to +1V, the control gate word lines are coupled to Vcc, the access gate word lines are coupled to Vcc, the source bit lines and the p-substrate node are coupled with 0V. The width of access device is made larger to lower device resistance in order to achieve a high speed read